DOCUMENT RESUME

ED 160 634 TM 007 715

AUTHOR Shell, Kevin D.; LeBold, William K.
TITLE The Utility of Multiple Discriminant Analysis in Classifying Students.
PUB DATE Mar 78
NOTE 22p.; Paper presented at the Annual Meeting of the American Educational Research Association (62nd, Toronto, Ontario, Canada, March 27-31, 1978)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS *Career Choice; College Freshmen; *Discriminant Analysis; *Engineering; Engineering Education; Engineers; Higher Education; *Interest Tests; *Predictive Validity; Specialization; *Vocational Interests
IDENTIFIERS Purdue Interest Questionnaire

ABSTRACT

The utility of multiple discriminant analysis in predicting career interests in six specialities of engineering--aeronautical, chemical, civil, electrical, industrial and mechanical engineering--was studied. Freshman engineers at Purdue University were administered the Purdue Interest Questionnaire (PIQ). Multiple discriminant analysis was performed on the six Engineering Specialty Scales of the PIQ using the obtained scores of 292 students. These students were then classified by means of discriminant functions into six engineering fields of study. Each predicted classification was then compared to the actual, fourth semester classification, and accuracy percentages for each field were computed. As a cross-validation of the discriminant analysis, these same discriminant functions were used to classify in the same manner a second sample of 301 students. Then the two student samples were pooled together, a new discriminant analysis was performed, and by means of the new functions the group of pooled students was again classified as before. In general, correct classification percentages significantly exceeded the percentages predicted by chance. Finally, it was observed in the cross-validation procedure that consideration by the counselor and a student of the first-or-second classification choice as optimal for success greatly increased the likelihood of making the best decision. (Author/CTM)

THE UTILITY OF MULTIPLE DISCRIMINANT ANALYSIS IN CLASSIFYING STUDENTS

by

Kevin D. Shell and William K. LeBold
Department of Freshman Engineering
Purdue University

Paper Presented at the Annual Meeting
American Educational Research Association
Division D

Toronto, Ontario, Canada
March 1978

(Printed in U.S.A.)

Session No. 24.14

THE UTILITY OF MULTIPLE DISCRIMINANT ANALYSIS IN CLASSIFYING STUDENTS

Kevin D. Shell and William K. LeBold
Purdue University

Abstract

This study examines the utility of multiple discriminant analysis in predicting engineering field of study. Representing six fields of study, 292 1970 and 301 1972 freshman engineers at Purdue University were administered the Purdue Interest Questionnaire (PIQ). Multiple discriminant analysis was performed on the six Engineering Specialty Scales of the PIQ using the obtained scores of the 1970 students. These students were then classified by means of discriminant functions into six engineering fields of study. Each predicted classification was then compared to the actual, fourth semester classification, and accuracy percentages for each field were computed. As a cross-validation of the discriminant analysis, these same discriminant functions were used to classify in the same manner the sample of 1972 students. Then the two student samples were pooled together, a new discriminant analysis was performed, and by means of the new functions the group of pooled students was again classified as before. In general, correct classification percentages significantly exceeded ($p < .001$) percentages predicted by chance. Finally, it was observed in the cross-validation procedure that consideration by the counselor and a student of the first-or-second classification choice as optimal for success greatly increases the likelihood of making the best decision.

The Utility of Multiple Discriminant Analysis in Classifying Students

Kevin D. Shell and William K. LeBold
Purdue University

The Department of Freshman Engineering at Purdue University endeavors to assist beginning students decide whether to pursue study in a special area of engineering or whether to consider a field of study outside engineering. For the most part, the students have been fairly certain in their desires to pursue a study of engineering. Therefore, the question of primary importance has been one of which specialty area of engineering would be most desirable to pursue.

Measures of ability and achievement, both during high school and during the first semester of college study, have found widespread use among counselors in assisting students through their academic endeavors. However, it also has been shown that a student's inventoried interests are a worthwhile source of additional information when considering appropriate areas of academic concentration within college. By far, the most widely utilized inventory measuring interests has in the past been the Strong Vocational Interest Blank (SVIB) -- now the Strong-Campbell Interest Inventory (SCII). All scales in the SVIB have been considered in examining the inventory's utility in differentiating special groups (Apostal, 1968; Chappel, 1967; Clemens & Linden, 1970). In particular, the engineering scale has been examined concerning its utility in predicting placement in engineering as a whole and in a specialty area within engineering (Curtis, 1970; Haddock, 1968; Mayfield, 1960; McCampbell, 1966). Of greater importance, however, has been the examination of the predictive utility of special scales, developed from the SVIB, in differentiating engineering specialty area (Haddock, 1968; Mayfield, 1960; McCampbell, 1966) and in predicting engineering achievement (Campbell, 1940; Curtis, 1970) and engineering persistence (Benjamin, 1967; Curtis, 1970; England, 1956; Haddock, 1968). A review of research concerning the utility of measured interests may be found in Benjamin (1967), Curtis (1970), DeLauretis (1975), and McCampbell (1966).

Efforts were begun in the mid 1960's to develop scales that would differentiate the major engineering specialty areas at Purdue University in order to assist the counselors within the Department of Freshman Engineering in their efforts to assist students. The fruit of these efforts was the Purdue Interest Questionnaire (PIQ) (LeBold and Gorman, 1966). The PIQ scales were analyzed and revised through the efforts of DeLauretis (1975), who examined the predictive utility of the revised scales in classifying both engineering students and transfer students from engineering. The current version of the PIQ is described and illustrated by LeBold, Shell, and DeLauretis (1977).

Although cross-validation studies of specialty scales, developed from the SVIB, have had encouraging results (Chappell, 1967; Curtis, 1970; McCampbell, 1966), no discriminant function cross-validation studies have been attempted with the PIQ. DeLauretis (1975) found that discriminant analysis using the PIQ scales produced discriminant functions that could accurately classify engineering students in their subsequent, fourth semester, specialty areas, but because of the unavailability of necessary data at that time he was not able to cross-validate the derived discriminant functions on an independent sample of engineering students. It was therefore one purpose of this study to perform a cross-validation of derived discriminant functions, using PIQ scales, in predicting the fourth semester classification of engineering students into their specialty areas.

Multiple discriminant analysis, in short, generates a set of weights which, when combined with the variable under investigation, maximize the individual differences <u>among</u> the classification groups and, at the same time, minimize the individual differences <u>within</u> each group. The rationale of multiple discriminant analysis is similar to that of multiple regression. However, in the case of classification group membership characterizes a nominal scale rather than an interval scale, which is necessary in justifying multiple regression procedures.

The procedures of multiple discriminant analysis may be examined in works by Cooley and Lohnes (1971), Rulon, Tiedeman, Tatsuoka, and Langmuir (1967), or Tatsuoka (1971). A summarized description can be found in DeLauretis (1975).

The purposes of this study were

1. to examine discriminant analysis results, using the six PIQ Engineering Specialty Scales, in predicting the fourth semester classification of engineering students into their specialty areas, including a cross-validation of derived discriminant functions;

2. to compare results of two classification procedures: minimum chi-square rule and maximum probability rule; and

3. to examine the predictive utility of considering first-or-second choices instead of only the first choice.

Methods

Subjects were sampled from freshman engineers enrolled in Purdue University's Department of Freshman Engineering in the fall semesters of 1970 (FE70) and 1972 (FE72). The FE70 sample consisted of 292 students, and the FE72 sample, of 301 students, consisting almost entirely of males. The FE70 sample represented six fields with samples from each field ranging from 29 to 85 students. The FE72 sample, representing the same six fields, contained between 21 and 85 students for each field. Each student was administered the Purdue Interest Questionnaire (PIQ), which contained thirteen scales, six of which concern specialty areas of study within engineering at Purdue: Aeronautical (AE), Chemical (CHE), Civil (CE), Electrical (EE), Industrial (IE), and Mechanical (ME). The raw score for each student on each scale was determined from response weights developed by DeLauretis (1975) in constructing the Purdue Engineering Interest Blank (PEIB). Standard scores were then computed based upon distributional statistics of a 1966 norm reference group, which consisted of 987 beginning freshman engineers.

The student scale scores are identical to those used by DeLauretis in constructing and examining the PEIB. These procedures were used in the present study because either they are still in current use with the PIQ or they afford the best means at present of utilizing the PIQ.

A major task for counselors within Freshman Engineering is the guidance of each beginning engineering student toward that area of engineering which best meets that student's abilities, interests, attitudes, etc. With this in mind one might tend to ask a question such as, "Which area of engineering should this particular student pursue?"

This type of question calls for the use of multiple discriminant analysis. A statistical routine (MDACAL, 1970), available at Purdue University's CDC 6500 computing facility[1], was used in performing the needed discriminant analyses and classifications. Two discriminant analyses were carried out with the six Engineering Specialty Scales, and discriminant functions were calculated for each analysis. First, the FE70 group was used in calculating discriminant functions, by means of which this FE70 group was reclassified (just as DeLauretis (1975) had previously done) and by means of which the FE72 group was independently classified, thus serving as a measure of the predictive power of the discriminant functions derived from the six scales. Second, the FE70 group and FE72 group were combined to form a single (hopefully more representative) group, new discriminant functions were calculated, and this combined group was then reclassified.

Each classification took place within the (reduced) discriminant space rather than in the original predictor space. Also, classification was performed according to each of two rules: the

---

[1] Adapted from Cooley, W.W. & Lohnes, P.R. Multiple Procedures for the Behavioral Sciences. 1962, 116-134.

minimum chi-square rule and the maximum (Bayesian) probability rule. According to the minimum chi-square rule, a student is classified into the category for which the distance between his discriminant score for that category and its corresponding centroid is the smallest. These distances follow a chi-square distribution. On the other hand, the maximum (Bayesian) probability rule classified the student according to the discriminant score having the greatest a priori probability of fit, i.e., following the proportional distribution of membership within each area of engineering study.

A "hit-miss" table was constructed for each classification, comparing the predictive utility of the minimum chi-square rule with that of the maximum probability rule. A "hit" occurs whenever predicted classification matches actual, fourth-semester classification. The overall percentage of hits, as well as the percentage of hits for each group were reported and served as the predictive power indices. In addition, during the independent classification (cross-validation of the discriminant functions) the overall percentage of first-or-second-choice hits, as well as the percentage of such hits for each engineering group, were calculated and compared with first-choice-only hits.

Each classification proportion was compared against the one-tailed critical value corresponding to the chance probability of classification, which was based upon the proportional group size with respect to the relevant-year group. The critical values were calculated by an estimation formula adapted from Hays (1973, p. 379):

$$p_{critical} = p_0 + z_{1-\alpha} \frac{\sqrt{p_0(1 - p_0)}}{\sqrt{n}}$$

where p_0 is the chance probability, z is the N(0,1) deviation score, α is the a priori statistical significance level, and n is the sample size. In addition, "pragmatically significant" differences -- those found to be consistent over classification attempts but not found to be statistically significant -- were examined.

Differences in hit proportions, as found by the two classification rules, were examined by means of the formula,

$$z = \frac{p_1 - p_2}{\sqrt{\frac{p_1(1 - p_1) + p_2(1 - p_2)}{n}}}$$

where p_1 is the hit proportion using the minimum chi-square rule, p_2 the hit proportion using the maximum probability rule; and n is the sample size (Downie & Heath, 1974, p. 182).

Results

As a guide to interpreting the classification results, Table 1 was constructed. The proportion of each engineering group to the total group was computed and identified as the probability that any student would be classified into that group by chance alone. From this a one-tailed critical value was estimated for each group and each classification procedure according to three levels of significance (α = .05; .01; .001) was entered as a percentage into Table 1. Thus each classification percentage can be compared with its relevant, corresponding (chance) critical value given in Table 1 as a means of statistically testing the predictive accuracy of the discriminant functions.

Tables 2, 3, and 4 show the accuracy with which students within each of the six engineering specialty groups were classified by the six-group discriminant analysis design. It is noted that the first entry within each cell of the tables is the percentage of students classified by the minimum chi-square rule, while the second entry is the percentage classified by the maximum (Bayesian) probability rule. Also note that diagonal cells display the percentage of each actual group being correctly classified.

Discriminant analysis was carried out on the six FE70 groups. Using the five discriminant functions calculated, the FE70 groups were reclassified (Table 2), and the FE72 groups were then independently classified (Table 3) as a check of the predictive utility

TABLE 1

APPROXIMATE ONE-TAILED CRITICAL VALUES[a] FOR CHANCE PROBABILITY OF CLASSIFICATION BASED UPON PROPORTIONAL GROUP SIZE

FE GROUP	α	FOR PREDICTED ENGINEERING GROUP					
		AE	CHE	CE	EE	IE	ME
1970	.05	21.3	21.7	25.2	37.2	19.1	28.2
	.01	25.1	25.5	28.9	40.6	22.8	31.7
	.001	29.2	29.6	32.9	44.3	27.0	35.7
1972	.05	20.3	18.4	27.5	36.3	16.2	33.4
	.01	24.0	22.1	31.0	39.6	20.0	36.8
	.001	28.1	26.3	34.9	43.3	24.2	40.6
1970 & 1972	.05	18.2	17.4	23.8	34.4	14.9	28.4
	.01	20.8	20.0	26.4	36.8	17.5	30.8
	.001	23.7	23.0	29.2	39.4	20.5	33.6

Note. n(FE70) = 292; n(FE72) = 301; n(FE70 & FE72) = 593.

[a] $p_{critical} = p_0 + z_{1-\alpha} \dfrac{\sqrt{p_0(1 - p_0)}}{\sqrt{n}}$, where n = sample size.

(or accuracy) of the five discriminant functions across samples. Subsequently, the FE70 groups and the FE72 groups were pooled, a new discriminant analysis was carried out, and by means of the five, new discriminant functions a reclassification was made of the pooled FE70 and FE72 groups (Table 4). Table 5 presents a summary of each engineering group's correct classification percentage as each of these three classifications was performed.

Tables 2 through 4 show all entries within the diagonal classification cells are significantly greater than their relevant, corresponding (chance) critical value given in Table 1 at the $p < .05$ level of significance (and for most cases beyond the $p < .001$ level) when the minimum chi-square rule is used. This also is true with but two exceptions (classification results

TABLE 2

RECLASSIFICATION MATRIX (SIX-GROUP DISCRIMINANT DESIGN): 1970 ENGINEERING SPECIALTY GROUPS (IN PERCENTAGES)

ACTUAL GROUP	PREDICTED GROUP					
	AE	CHE	CE	EE	IE	ME
AE (n = 36)	50.0[3] 50.0[3]	11.1 8.3	2.8 8.3	11.1 19.4	---- ----	25.0* 13.9
CHE (n = 37)	2.7 2.7	78.4[3] 73.0[3]	---- 2.7	8.1 13.5	2.7 2.7	8.1 5.4
CE (n = 48)	---- ----	---- ----	75.0[3] 85.4[3]	---- 2.1	12.5* 6.2	12.5 6.2
EE (n = 85)	4.7 3.5	7.1 5.9	2.4 3.5	70.6[3] 80.0[3]	---- ----	15.3 7.1
IE (n = 29)	3.4 6.9	6.9 6.9	10.3 13.8	---- 3.4	65.5[3] 62.1[3]	13.8 6.9
ME (n = 57)	14.0* 17.5*	5.3 3.5	7.0 14.0	12.3 15.8	8.8 8.8	52.6[3] 40.4[3]
OVERALL PERCENTAGE CORRECTLY CLASSIFIED (n = 292)						65.8 66.8

Note 1. The first entry in each cell follows the minimum chi-square rule. The second entry follows the maximum probability rule. n is the number of students within the actual group.

Note 2. Percentages are compared to critical values for 1970 group in Table 1 for significance.

*Above chance probability but not significant

[1] Significant at $p < .05$

[2] Significant at $p < .01$

[3] Significant at $p < .001$

TABLE 3

INDEPENDENT CLASSIFICATION MATRIX (SIX-GROUP DISCRIMINANT DESIGN): FE72 ENGINEERING SPECIALTY GROUPS CLASSIFIED FROM FE70 ENGINEERING SPECIALTY GROUP DISCRIMINANT FUNCTIONS (IN PERCENTAGES)

ACTUAL GROUP	PREDICTED GROUP					
	AE	CHE	CE	EE	IE	ME
AE (n = 34)	35.3^{3} 29.4^{3}	5.9 5.9	8.8 11.8	17.6 26.5	2.9 2.9	29.4* 23.5
CHE (n = 28)	3.6 7.1	75.0^{3} 64.3^{3}	---- 3.6	14.3 21.4	3.6 ----	3.6 3.6
CE (n = 57)	8.8 8.8	10.5* 7.0	38.6^{3} 47.4^{3}	10.5 13.3	1.8 1.8	29.8* 22.8
EE (n = 85)	3.5 2.4	11.8* 7.1	2.4 2.4	67.1^{3} 76.5^{3}	1.2 1.2	14.1 10.6
IE (n = 21)	---- ----	4.8 ----	9.5 23.8^{2}	---- 14.3	33.3^{3} 19.0^{1}	52.4^{3} 42.9^{3}
ME (n = 76)	19.7* 19.7*	14.5* 11.8*	17.1 18.4	10.5 17.1	2.6 1.3	35.5^{1} 31.6*
OVERALL PERCENTAGE CORRECTLY CLASSIFIED (n = 301)						48.5 49.2

Note 1. The first entry in each cell follows the minimum chi-square rule. The second entry follows the maximum probability rule.

Note 2. Percentages are compared to critical values for 1972 group in Table 1 for significance.

*Above chance probability but not significant

[1] Significant at $p < .05$

[2] Significant at $p < .01$

[3] Significant at $p < .001$

TABLE 4

RECLASSIFICATION MATRIX (SIX-GROUP DISCRIMINANT DESIGN): FE70 & FE72 ENGINEERING SPECIALTY GROUPS COMBINED (IN PERCENTAGES)

ACTUAL GROUP	PREDICTED GROUP AE	CHE	CE	EE	IE	ME
AE	42.9[3]	8.6	7.1	10.0	1.4	30.0[1]
(n = 70)	38.6[3]	8.6	11.4	20.0	1.4	20.0
CHE	1.5	75.4[3]	3.1	9.2	4.6	6.2
(n = 65)	3.1	64.6[3]	6.2	15.4	4.6	6.2
CE	2.9	4.8	60.0[3]	5.7	6.7	20.0
(n = 105)	2.9	1.9	67.6[3]	10.5	3.8	13.3
EE	4.7	9.4	3.5	65.9[3]	2.4	14.1
(n = 170)	2.9	4.1	3.5	79.4[3]	1.8	8.2
IE	6.0	4.0	10.0	4.0	64.0[3]	12.0
(n = 50)	6.0	2.0	12.0	8.0	60.0[3]	12.0
ME	17.3*	8.3	16.5	8.3	9.8*	39.8[3]
(n = 133)	16.5*	6.0	24.1[1]	18.8	6.8*	27.8*
OVERALL PERCENTAGE CORRECTLY CLASSIFIED						57.2
(n = 593)						57.7

Note 1. The first entry in each cell follows the minimum chi-square rule. The second entry follows the maximum probability rule.

Note 2. Percentages are compared to critical values for 1970 and 1972 group in Table 1 for significance.

*Above chance probability but not significant

[1] Significant at $p < .05$

[2] Significant at $p < .01$

[3] Significant at $p < .001$

regarding Mechanical Engineering (ME students given in Tables 3 and 4) when the maximum probability rule is used. Thus, all six groups of students were in general classified correctly to a significant degree above chance. As an example, consider the Civil Engineers (CE) of the FE72 engineering specialty groups, classified by the discriminant functions calculated from the FE70 engineering specialty groups (Table 3). In examining Table 1, it is seen that for the FE72 groups the chance critical values corresponding to the CE group for the $p < .05$, 01, and .001 levels of significance are respectively 27.5, 31.0, and 34.9. As Table 3 shows, the minimum chi-square rule results in a correct classification percentage of 38.6, while the maximum probability rule leads to 47.4 percent being correctly classified. Both these percentages are above the $p < .001$ critical value of 34.9.

In further examining Tables 2 through 4, it is seen that Aeronautical Engineers (AE) were consistently misclassified across the three classification procedures (i.e., for FE70, for FE72, and for pooled FE70 and FE72) above chance as Mechanical Engineers (ME)(significant at $p < .05$ for the pooled group of Table 4, i.e., 30.0 > 28.4). This was reciprocal in that with both classification rules ME students were consistently (although not significantly) misclassified above chance across the three classifications as AE students.

It should also be noted that in the FE72 independent classification Industrial Engineers (IE) were not only significantly misclassified above chance at $p < .001$ as ME students but were misclassified (by both rules) to an extent well above the percentage of correct classification (i.e., 52.4 > 33.3 and 42.9 >19.0).

In the examination of Table 5 it is seen that the overall percentage of correct group classification is rather high for all three classification procedures using both classification rules. Although the maximum probability rule consistently produced better results that the minimum chi-square rule, in general

TABLE 5

CORRECT CLASSIFICATION
PERCENTAGE MATRIX:
ENGINEERING SPECIALTY GROUPS
(SIX-GROUP DISCRIMINANT DESIGN)

ACTUAL GROUP	FE70	FE72	FE70 & FE72
AE	50.0 50.0	35.3 29.4	42.9 38.6
CHE	78.4 73.0	75.0 64.3	75.4 64.6
CE	75.0 85.4	38.6 47.4	60.0 67.6
EE	70.6 80.0	67.1 76.5	65.9[2] 79.4
IE	65.5 62.1	33.3 19.0	64.0 60.0
ME	52.6 40.4	35.5 31.6	39.8[1] 27.8
OVERALL PERCENT-AGE	65.8 66.8	48.5 49.2	57.2 57.7

Note. The first entry in each cell follows the minimum chi-square rule. The second entry follows the maximum probability rule.

[1]Difference significant at $p < .05$

[2]Difference significant at $p < .01$

there was no significant difference between the outcomes of the respective rules.

During the independent classification procedure not only were percentages calculated of first-choice correct classifications ("hits"), but also calculated were the percentages of correct classifications with respect to the first-or-second-choice--correct second-choice classifications are sometimes referred to in the literature as "near hits." These results are presented in Table 6. However, momentarily returning to Table 5, it is seen that, in proceeding to the FE72 independent classification from the FE70 reclassification, a large decrease in correct classifications was observed for some groups (e.g., AE, CE, IE, and ME). However, as is seen in Table 6, by considering "near hits" as well as "hits," the correct classifications increased substantially although not equally across groups (e.g., the increase in the CE group as compared with the increase in the other groups). The overall percentage correctly

TABLE 6

CORRECT INDEPENDENT-CLASSIFICATION PERCENTAGE MATRIX:
(SIX-GROUP DISCRIMINANT DESIGN)
ENGINEERING SPECIALTY SCALES

(FE72 CLASSIFIED FROM FE70 DISCRIMINANT FUNCTIONS)

ACTUAL GROUP	CORRECT CLASSIFICATION PERCENTAGE	
	1ST CHOICE	1ST OR 2ND CHOICE
AE (n = 34)	35.3 29.4	79.4 76.5
CHE (n = 28)	75.0 64.3	82.1 82.1
CE (n = 57)	38.6 47.4	56.1 59.6
EE (n = 85)	67.1 76.5	80.0 83.5
IE (n = 21)	33.3 19.0	52.4 42.9
ME (n = 76)	35.5 31.6	65.8 68.4
OVERALL (n = 301)	48.5 49.2	70.1 71.4

Note. The first entry in each cell follows the minimum chi-square rule. The second entry follows the maximum probability rule.

classified by the minimum chi-square rule increases from 48.5 to 70.1 while that by the maximum probability rule increases from 49.2 to 71.4.

Discussion

The results show that engineering students can be classified accurately using functions derived from a discriminant analysis of the six PIQ Engineering Specialty Scales. The classifications were found in general to be significantly better than chance predictions at the $p < .001$ significance level. This significance remained during cross-validation of the discriminant functions even though percentages of correct classification did decrease and, for most specialty groups, substantitally.[2] These classification results not only show good utility of the derived discriminant functions but also good utility of the six PIQ Engineering Specialty Scales used in deriving the discriminant functions.

Nevertheless, problems do exist for some of the scales. As observed in the results, Aeronautical Engineers and Mechanical Engineers were at times confounded with each other. It was also observed that with the independent, FE72 cross-validation group IE students were significantly misclassified (above $p < .001$) above chance as ME students. In fact, a larger percentage of IE students were misclassified than were correctly classified. It should be noted that the AE and IE shcools were at one time incorporated with the ME school. The classification confusion seems to indicate that characteristics unique to the FE70 sample of IE and ME (and possibly AE)students may have been present and thus adversely affected the discriminant functions, thus calling for re-examination and revision of their relevant scales rather than a deficiency in the discriminant analysis. The confounding may also have been due in part to somewhat small sample sizes. An increase in the sample size of each engineering group (currently in progress) may thus alleviate somewhat the misclassification tendencies by making each sample group more representative of its own population and thereby lessening any unique characteristic effects. Efforts are also underway to increase the number of items per scale as well as the number of scales.

---

2 This decrease in discriminant function predictability may be compared conceptually to the decrease in multiple regression predictability (shrinkage of the multiple R^2 or variance accounted for) over samples.

Although all groups correctly classified above chance to a significant degree, deficiencies are seen when the percentages of correct classifications are compared with their corresponding percentages of total misclassifications. Since for several groups more students were misclassified than were correctly classified, there may be important characteristics of each group remaining to be measured. Also, of possible relevance to this is the discussion following.

When both the first and the second choice for classification were considered (in the cross-validation), a large increment (relative to the increment possible) was observed in the percentage of correct classifications. This would indicate that a counselor should present to students for their consideration the first two choices as optimal ones and use other information available, such as abilities and stated preferences, to help decide between the two choices. This could even be extended to the top three (out of six) choices as a means of effectively reducing the number of choices that must be considered. A feasible alternative would be to actually use scores from relevant ability measures in the discriminant analysis along with interest scores. Too often a counselor -- admittedly not the best counselors -- will present only "the best choice" possible (as the counselor sees it) or will present all possible choices with little guidance toward the best or most optimal choices (as determined by the student's objectives, abilities, interests, etc.). The proposed plan above should hopefully help students feel less restricted in their endeavors and at the same time help them pursue a realistic career alternative.

According to the above methodology, discriminant analysis is conducted with p groups and the top g groups are selected as the most optimal for an individual -- possessing the closest match between the individual's scores and group characteristics. This methodology is conducive to a counseling viewpoint which concerns examining p possible groups together and selecting the g most likely choices. However, an alternate viewpoint involves examining group similarity as a yes/no question (with a possible criterion) for an individual and each possible group. According to this viewpoint p separate, two-group discriminant analyses would be conducted involving membership versus non-membership (individuals in one group versus all other individuals) for each particular group. Probability of the individual's membership would be calculated for each of the p groups. The counselor could then select those groups whose probabilities surpass a criterion or which have the highest probabilities -- if none of the probabilities surpass the criterion. DeLauretis (1975) examined two-group discriminant analysis functions in reclassifying each of the six FE70 engineering groups versus a group of freshman engineers in general and obtained greater results for each group (percentage of correct classifications greater than 75 percent with median percentage of 84.6) than during the FE70 six-group discriminant analysis and reclassification. It should be noted that these analyses conducted by DeLauretis are only similar to that proposed above since the reference group was a group of freshman engineers in general rather than a composite of other groups of upperclass engineers. Nevertheless, this gives some encouraging credibility to the two-group discriminant analysis approach, although he was not able at that time to examine the cross-validation of his two-group discriminant functions on the FE72 sample.

The several, two-group discriminant analysis approach may be compared to performing several, separate t-tests rather than an all-encompassing analysis of variance. Thus, it may be that th[illegible]ach may not only greatly restrict de[illegible]s of freedom for infer[illegible] may also increase the likelihood of matching an individual to a [illegible]milar group. However, considering DeLauretis' results, it may be worthwhile to compare the predictive utilities of these two discriminant app[illegible]aches.

Although the maximum (Bayesian) probability rule was observed to be consistently superior to the minimum chi-square rule in the overall percentage of correct classifications, the difference was not significant, and the superiority did not hold across groups. It is thought that any "superiority" may have been due, at least in part, to the unequal group sample sizes, but this is not completely substantiated by the data. There is therefore a need to further examine this possible tendency. It is also suggested that the use of one rule with some groups and the other rule with the remaining groups may be considered and further studied.

Conclusions

Discriminant analysis functions, based on the six PIQ Engineering Specialty Scales, correctly classified students into their specialty group significantly above chance predictions ($p < .001$, in general). Although the percentages of correct classification decreased during cross-validation of the discriminant functions, the percentages were in general significantly above chance levels. Therefore, discriminant analysis (as well as the six PIQ Engineering Specialty Scales) is considered to be <u>of assistance</u> to beginning engineering students and counselors in making decisions concerning the engineering specialties. It may also prove worthwhile to consider not only the first, but also the second choice of specialty area as being optimal for success and then to use other information available to make the most reasonable decision.

Although problems were observed, indicating that a re-examination of each scale may be necessary in or [illegible] es more char- acteristic of their respective [illegible] remains important note that, even with these proble[illegible] e scales and discriminant functions show good utility in adding [illegible] the available information about a student in order to assist in decisions concerning future academic endeavors.

Educational Impact

Engineering is one of the largest professions in the United States today with over a quarter million undergraduate students (over one million practicing engineers in 1972). However, only about one-half of the students who begin engineering study actually graduate in an engineering field, and many who remain in engineering change specialty area within engineering prior to graduation. In addition the engineering profession is becoming more complex and broadening its responsibilities and goals. It therefore becomes increasingly important to guide capable and interested individuals toward meeting the need of this expanding profession and of the nation. It has been shown that inventoried interests (such as measured by the PIQ) provide worthwhile information and multiple discriminant analysis a worthwhile method of using this information for assisting with educational/career decisions. However, it is important to note that these techniques need -- and should -- not be limited to engineering but could be applied to the broader domain of career fields available and could be very useful as occupational interests within individuals develop.

References

Apostal, R.A. Interests of engineering graduates according to undergraduate curricula. Personnel and Guidance Journal, 1968, 46, 909-913.

Benjamin, D.R. A th[illegible] one year longitudinal study of engineering students interest profiles and career patterns. (Doctoral dissertation, Purdue University, 1967) Dissertation Abstracts International, 1968, 28(11), 4441A. (University Microfilms No. 68-06281)

Campbell, D.P. Manual for the Strong-Campbell Interest Inventory, Second Edition. Stanford, Calif.: Stanford University Press, 1977.

Chappell, J.S. Multivariate discrimination among selected occupational groups utilizing self report data. (Doctoral dissertation, Purdue University, 1967) Dissertation Abstracts International, 1967, 28, 2620B. (University Microfilms No. 67-16625)

Clemens, B. & Linden, J. Discriminant function analysis of inventoried interests among selected engineering groups. Purdue University, March 1970. (A paper presented at the Annual Meeting of the American Educational Research Association) (ED 041 945)

Cooley, W.W. & Lohnes, P.R. Multivariate Data Analysis. New York: John Wiley & Sons, 1971.

Curtis, J.T. Inventoried interests as related to persistence and academic achievement in an engineering program. (Doctoral dissertation, Purdue University, 1970) Dissertation Abstracts International, 1970, 31, 1572A-1573A. (University Microfilms No. 70-18623)

DeLauretis, R.J. The construction of the Purdue Engineering Interest Blank: a unit-scaling and optimal-scaling approach. (Doctoral dissertation, Purdue University, 1975) (University Microfilms No. 76-21093)

Downie, N.M. & Heath, R.W. Basic Statistical Methods, Fourth Edition. New York: Harper & Row, 1974.

England, G.W. The interest factor in undergraduate engineering achievement. (Doctoral dissertation, University of Minnesota, 1956) Dissertation Abstracts International, 1956, 17(4), 902B. (University Microfilms No. 00-19199)

Haddock, R.A. Differentiation of engineering students' interests. (Unpublished masters thesis, Purdue University, 1968)

Hays, W.L. Statistics for the Social Sciences, Second Edition. New York: Holt, Rinehart, & Winston, 1973.

LeBold, W.K. & Gorman, W. Purdue Interest Questionnaire. 1966 (Current version: Purdue Research Foundation, 1976. Available from W.K. LeBold, Department of Freshman Engineering, Purdue University, West Lafayette, Indiana 47907)

LeBold, W.K., Shell, K.D., & DeLauretis, R. The Purdue Interest Questionnaire -- an interest inventory to assist engineering students in their career planning. In L.P. Grayson & J.M. Biedenbach (Eds.), Proceedings, Seventh Annual Frontiers in Education Conference, The Institute of Electrical and Electrical-and-Electronic Engineers, Inc. & American Society for Engineering Education, 1977, 488-495.

Mayfield, E.C. Interest as a predictor of graduation in engineering. (Doctoral dissertation, Purdue University, 1960) (University Microfilms No. 60-4188)

McCampbell, M.K. Differentiation of engineers' interests. (Doctoral dissertation, University of Kansas, 1966) Dissertation Abstracts, 1966, 27, 1662A-1663A. (University Microfilms No. 66-13051)

Rulon, P.J., Tiedeman, D.V., Tatsuoka, M.M., & Langmuir, C.R. Multivariate Statistics for Personnel Classification. New York: John Wiley & Sons, 1967.

Tatsuoka, M.M. Multivariate Analysis: Techniques for Educational and Psychological Research. New York: John Wiley & Sons, 1971.